Proc. Natl. Acad. Sci. USA
Vol. 81, pp. 974–978, February 1984
Physiological Sciences

Quantitative investigation of hepatic genomic response to hormonal and pathophysiological stimuli by multivariate analysis of two-dimensional mRNA activity profiles

(translation products/quantitation/physiological interrelationships/statistical analysis)

F. E. CARR*, C. BINGHAM†, J. H. OPPENHEIMER*, C. KISTNER†, AND C. N. MARIASH*

*Department of Medicine, University of Minnesota, Minneapolis, MN 55455; and †Department of Applied Statistics, University of Minnesota, St. Paul, MN 55108

Communicated by Ralph T. Holman, October 31, 1983

ABSTRACT **We have applied techniques of multivariate analysis to the characterization and comparison of the effects of various pathophysiological and hormonal stimuli on the expression of the rat hepatic genome at a pretranslational level. *In vitro* translated products were resolved by two-dimensional gel electrophoresis. We analyzed 10 pathophysiological states brought about by variation in thyroidal status, starvation, administration of high carbohydrate diet, and the production of experimentally induced diabetes mellitus. Each state differed significantly from every other state in the two-dimensional electrophoretic profiles. The set consisting of the minimal number of products necessary for maintaining the distinctive patterns was identified. The analysis also defined those clusters of products that behaved in a coordinate fashion in response to the various stimuli. Lastly, the similarity and dissimilarity of hepatic mRNA activity profiles to each other could be geometrically represented in three-dimensional space. Our finding that the hepatic mRNA activity profile could distinguish reliably between closely related hormonal and pathophysiological stimuli indicates the specificity of hepatic genomic expression. A systematic analysis of such profiles may be useful as an overall index of the biologic response at the hepatocellular level.**

The ability to separate by two-dimensional gel electrophoresis the *in vitro* translational products of total cytoplasmic poly(A)-containing RNA has provided a much broader range of information than previously available regarding the genomic response to various hormonal and pathophysiological stimuli (1–4). At the same time, computerized videodensitometry has enormously facilitated the acquisition, quantitation, and storage of such information (5). These techniques have allowed us to compare the genomic response elicited by hypothyroidism and starvation by contrasting the directional shifts of individual mRNA sequences (4). However, a more comprehensive analytic approach appeared to be desirable in order to provide a precise assessment of the overall degree of similarity or dissimilarity of the mRNA activity profile patterns that characterize specific pathophysiological and hormonal states. Moreover, given the inherent biological and experimental variation of the system, it appeared important to provide an objective method to determine whether profiles from any two given states are statistically distinct. We also wished to determine how few products were required to separate individual states in any given set of states and to identify those clusters of products that change in a coordinate fashion in the various transitions studied. The statistical techniques of multivariate analysis appeared ideally suited to achieve these objectives because each gel could provide a single multivariable observation whose components were the quantitated product densities on that gel.

MATERIALS AND METHODS

Animal Treatments. Male Sprague–Dawley rats (175–250 g) were supplied by Taconic Farms, (Germantown, NY). Pathophysiological and hormonal stimuli were applied as described in Table 1. We shall refer to these states by the labels (TX, EU, . . .) defined there. The specific stimuli were selected in accordance with the overall interests of this laboratory in the nature of the thyroid hormone–carbohydrate interaction. Data from groups 1–4 (Table 1) have been used in previous studies (4). Translational assays were performed as described (3, 6). RNA was extracted from frozen liver with phenol/chloroform, 1:1 (vol/vol). Poly(A)-containing RNA was subsequently recovered by oligo(dT)-cellulose chromatography and translated in an *in vitro* endonuclease-treated reticulocyte lysate system containing [^{35}S]methionine. Translated products were resolved by two-dimensional gel electrophoresis according to the methods of O'Farrell (8). Individual spots were quantitated by computerized videodensitometry as developed by Mariash *et al.* (5). Each gel used in the final data analysis was obtained from a single animal and represented the results of a single translational assay. Frozen liver in these studies were thawed only once.

Each step of the analysis is subject to unavoidable experimental variation. Thus, RNA extraction of specific mRNAs may vary from specimen to specimen. There also may be differential recovery of specific mRNAs from oligo(dT) columns, and the action of ribonuclease may exert variable effects during extraction, storage, and analysis. Despite these potential sources of error, rigorous standardzation of the procedure has allowed us to provide satisfactory replicate determinations in the same specimen and allowed us to draw physiological and biochemical inferences about the genomic effects of various pathophysiological stimuli (3, 5, 9, 10, ‡). Moreover, in recent studies, we have shown an excellent correlation between hybridization analysis of spot 14 and translational assays in our laboratory.‡ Nevertheless, potential random variation due to uncontrolled technical factors must be taken into account in any analysis of the results. The statistical approaches used in this study encompass both the experimental and the biological sources of variation.

Statistical Analysis. Visual inspection of 50 gels, including at least 2 from each of the 10 states, allowed selection of a number of spots that appeared to vary from one state to another. These spots were quantitated by computerized videodensitometry (5). On the basis of two-sample Student's *t* sta-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: T3, triiodothyronine.

‡Jump, D. B., Narayan, P., Towle, H. C. & Oppenheimer, J. H. (1983) Program of the Endocrine Society, 65th Annual Meeting, June 8–10, San Antonio, Texas, Abstract #21.

Table 1. Animal treatments*

State	Label	*N*	Treatment
1 Hypothyroid	TX	4	Surgically thyroidectomized; 1 wk of low-iodine diet; treated with 100 μCi of ^{131}I.
2 Euthyroid	EU	11	Euthyroid rats received standard rat chow and water ad lib (control).
3 Fasted	FAST	7	Euthyroid rats were deprived of food for 5 days.
4 Fasted + T3	FAST + T3	3	Euthyroid rats were deprived of food for 5 days, injected i.p. with a single receptor-saturating dose of T3 (200 μg/100 g of BW) on day 4.
5 Moderate diabetes	DB − M	7	Euthyroid rats injected with streptozotocin (50 mg/kg of BW) followed by a single i.p. injection of 2 ml of 30% glucose to minimize transient hypoglycemia; animals were killed 10 days later; fasting plasma glucose, >200 mg %.
6 Moderate diabetes + insulin	DB − M + I	3	Euthyroid rats rendered diabetic as above; treated with pork insulin (2.4 units/day delivered at 1 μl/hr from Alzet pump placed subcutaneously) for additional 7 days.
7 Severe diabetes	DB − S	2	Euthyroid rats injected with streptozotocin (75 mg/kg of BW) followed by a single i.p. injection of 2 ml of 30% glucose to minimize transient hypoglycemia; animals were killed 4 days later; fasting plasma glucose, >350 mg %.
8 High carbohydrate diet	HiCHO	5	Euthyroid rats fed diet of 20% protein/60% sucrose/20% bulk for 10 days.
9 Euthyroid + T3†	EU + T3	4	Euthyroid rats treated with a single i.p. injection of T3 (200 μg/100 g of BW) 24 hr before sacrifice.
10 Hyperthyroid†	HYPER	4	Euthyroid rats injected i.p. with T3 (15 μg/100 g of BW) for 10 days.

BW, body weight.

*Animals received food and water ad lib unless otherwise indicated.

†Euthyroid rats treated with a receptor-saturating dose of T3 (200 μg/100 g of BW) are metabolically distinct from hyperthyroid rats. In part, this is a reflection of various half-lives of responsive enzymes. As previously shown, malic enzyme activity increases from euthyroid levels of 12.9 ± 3.8 to 47.6 ± 10.2 units/mg of protein with T3 treatment (200 μg/100 g of BW) for 1 day (7) and to 141 ± 7 units/mg of protein in hyperthyroid animals (6).

tistics, a set of 42 spots was selected, for each of which at least two states were significantly different ($P < 0.05$). To determine the effectiveness of the visual inspection, additional products were randomly selected and quantitated. None displayed significantly changed levels between any two states. Not all of the 42 products, however, could be reliably quantitated for each gel. The resultant incomplete data presented a major problem to the application of multivariate analysis. In the subsequent analysis, we therefore restricted our analysis to 19 translated products with no missing data on any of the 50 gels representing the 10 physiological states. The translated products are numbered in accordance with previous studies from our laboratory (3, 4). The cpm of individual products was transformed by the function ln(cpm + 1) in order to minimize the systematic dependence of variability on the magnitude of the mean value. Rankit plots (11) of residuals were made as an informal check of normality.

To determine statistical significance of differences between states, pair-wise comparisons of the means were made using Hotelling's T^2 based on all 19 products and using univariate two-sample statistics for each product separately (12, 13).

Stepwise discriminant analysis[§] was used to identify the variables (translated products) that best distinguished all 10 states as separate groups. This procedure further facilitated the selection of a small number of products that allowed all gels to be assigned to states with a small probability of error.

Cluster analysis (14, ¶) is a technique of multivariate analysis that allows segregation of a collection of "objects" into subsets of "clusters" such that the objects in a cluster are more similar to each other than to objects not in the cluster. Cluster analysis was applied to two distinct problems. In the first, the objects clustered were the 50 individual two-dimensional gel patterns. Primary clusters were formed by grouping together sets of gels with the least dissimilar patterns. As the requirements for dissimilarity increased, the primary clusters merged into larger clusters. This process was continued with progressively larger clusters merging until all the constituent gel patterns had been incorporated into one all-inclusive group. If a state can be characterized, in fact, by a distinct mRNA activity profile, then gels from that state should form a primary cluster, and, as the analysis proceeds, states with similar patterns should merge to form larger clusters. In the second problem, the objects clustered were the 19 products. Similarity was determined with respect to their behavior in the various transitions under study. Products exhibiting similar patterns of changes among the 10 states were grouped into a primary cluster. With progressively increasing requirements for dissimilarity, the primary clusters merged to form larger groups of products. For both types of clustering, the single-linkage algorithm was used.

Finally, by using the 19 products simultaneously, the relative similarity of all the states was assessed by an analysis of canonical variables derived from a one-way multivariate analysis of variance (15, ‖). This technique provides a quantitative assessment of the degree of similarity of each of the 10 states to each of the other states. Because in our analysis each state is determined by 19 variables, the individual products, the state mean can be considered as a point in a 19-

[§]BMDP, Biomedical Computer Programs, P-Series, Health Science Computing Facility, AV-111, CHS, UCLA, CA 90024.

¶Loesch, J. H. CLUSTER: A Cluster Analyses Package, Version 1.0., Social Sciences Research Facilities Center, University of Minnesota, Minneapolis, MN.

‖Bingham, C. MATTER: A Program for Interactive Matrix and Time Series Computation, Department of Applied Statistics, University of Minnesota, St. Paul, MN.

Table 2. Translated products analyzed for different pathophysiological states

	State									
Product	EU	TX	FAST	FAST + T_3	DB − M	DBM + I	DB − S	HiCHO	EU + T_3	HYPER
3	243.4	0.00	171.0	106	148.7	24.0	194.5	118.6	198.2	281.5
17	547.1	905.0	1096.4	672.3	1175.9	536.0	1322.0	351.2	242.3	334.0
29	222.1	351.25	347.43	381.0	339.3	424.7	346.0	45.2	320.2	108.0
33	0.27	1.71	0	0	22.1	4.7	42.0	1.80	0	0
14	118.4	29.0	4.0	228.7	5.6	3.0	0.5	509.6	109.2	639.3
10	655.8	519.0	486.1	80.3	514.4	405.0	1111.5	71.6	163.7	118.5
12	154.5	10.25	111.6	192.0	101.4	187.3	105.0	77.0	162.7	168.0
4	70.9	4.75	60.7	88.7	66.6	144.0	229.5	237.6	175.2	389.5
11	185.5	46.2	123.7	281.7	170.4	150.3	213	289.4	343.0	545.7
8	217.6	1.00	172.6	107.7	281.3	142.7	311.5	160.6	228.7	138.2
15	2256.5	18.5	568.4	78.7	477.7	100.0	306.5	533.4	1012.0	1266.0
5	46.4	3.25	18.7	39.0	22.0	30.3	68.5	83.0	68.7	134.5
41	167.9	404.7	276.4	190.0	242.1	266.0	222.0	175.2	181.5	160.5
24	124.3	340.7	404.9	147.0	314.0	193.7	532.0	69.0	46.7	129.8
25	78.9	155.2	242.6	104.3	233.4	95.7	186.0	27.8	33.7	2.7
13	19.5	45.5	32.9	19.3	18.6	15.7	20.5	11.6	35.0	8.3
37	84.1	120.5	68.4	42.7	63.7	48.3	161.0	36.4	52.3	84.5
31	51.2	56.0	73.1	98.3	56.0	22.0	79.5	27.2	52.0	56.25
40	204.7	186.0	163.7	180.3	128.6	134.3	134.0	266.6	236.2	223.7
Sample size	11	4	7	3	7	3	2	5	4	4

Data are expressed as the mean cpm values for the translated products. Equal cpm (200,000) were applied to each gel analyzed. The first 10 rows are the products selected by stepwise discriminant analysis to separate all states and are in the order they entered the discriminant functions.

dimensional space. The first canonical variable can be considered as the axis that provides maximal separation of all state means in the direction determined by that axis. The second canonical variable provides maximal separation in a direction perpendicular to the first. The third canonical variable provides maximal separation in a direction perpendicular to the plane determined by the first and second canonical variables. The fourth canonical variable is perpendicular to the space created by the first three, and so on. Because three canonical variables can explain 90% of the variance of the 19 dimensions, the relationships among the states can be geometrically represented in three dimensions. States with similar two-dimensional gel profiles should then be located near each other.

RESULTS

The 10 states differed among themselves significantly ($P < 0.001$) based on both the MANOVA maximum root test and separate univariate F tests. All pair-wise univariate t statistics showed that each pair of states is significantly different ($P < 0.05$) by the Tukey–Kramer multiple comparisons procedure.‡ However, by using a conservative form of Hotelling's T^2 analysis with Bonferroni's inequality (12), all pair-wise comparisons were significant ($P < 0.05$) except DB-S and DB-M ($P < 0.09$). The failure to achieve significance between these states is in part due to the small sample for DB-S ($N = 2$).

Without limitation in the number of variables selected, step-wise discriminant analysis selected 10 products (3, 17, 29, 33, 14, 10, 12, 4, 11, 8) (Table 2) in descending order of their ability to distinguish among all states. When the accuracy of discrimination among the states was evaluated with these 10 spots, 48 of 50 gels were correctly assigned to their associated states with 1 DB-M incorrectly assigned to DB-S and 1 DB-S assigned to DB-M. Discrimination based on the first four products (3, 17, 29, 33) resulted in 47 correct assignments.

Cluster analysis (Fig. 1) confirmed that the quantitated products from the gels could distinguish the different states solely on the basis of the pattern determined by the 19 products. All 50 gels were clustered ignoring the state with which each gel was associated. Overlap of primary clusters was noted only among the diabetic states. Fig. 1 also displays how clusters (groups of apparently similar gels) merge to form larger clusters. For example, gels derived from the fasting state merged with a cluster consisting of gels from the three diabetic states indicating similarity in the translated products affected. The hyperthyroid and high carbohydrate states merged but the hypothyroid state remained distinct.

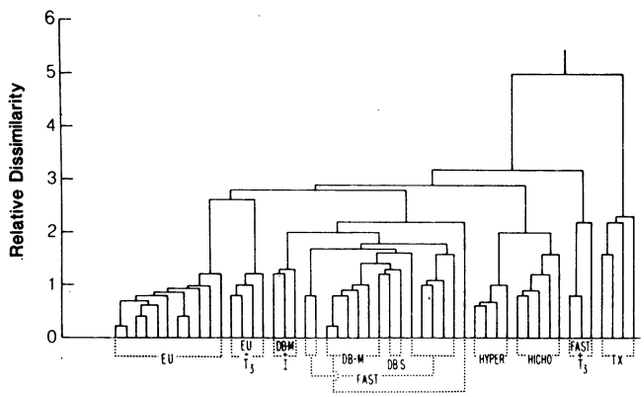

FIG. 1. Dendrograms representing a hierarchial cluster analysis of the 50 gels. Each "root" is associated with a gel derived from 1 of the 10 physiological states. Dissimilarity of gels was measured by euclidean distance based on 19 translated products. The single linkage (minimum distance) was used (13). All gels linked within a group are closer (more similar) to some other gel within that group than to any other gel outside. The height at which the linkage occurs reflects the relative dissimilarity, with most similar states linked first (lowest). Linkage progresses until all states are contained within a single group. The units indicated are of relative distance from the point where all 50 gel states are distinct (0) to where all gels and, hence, all states are within one group (5).

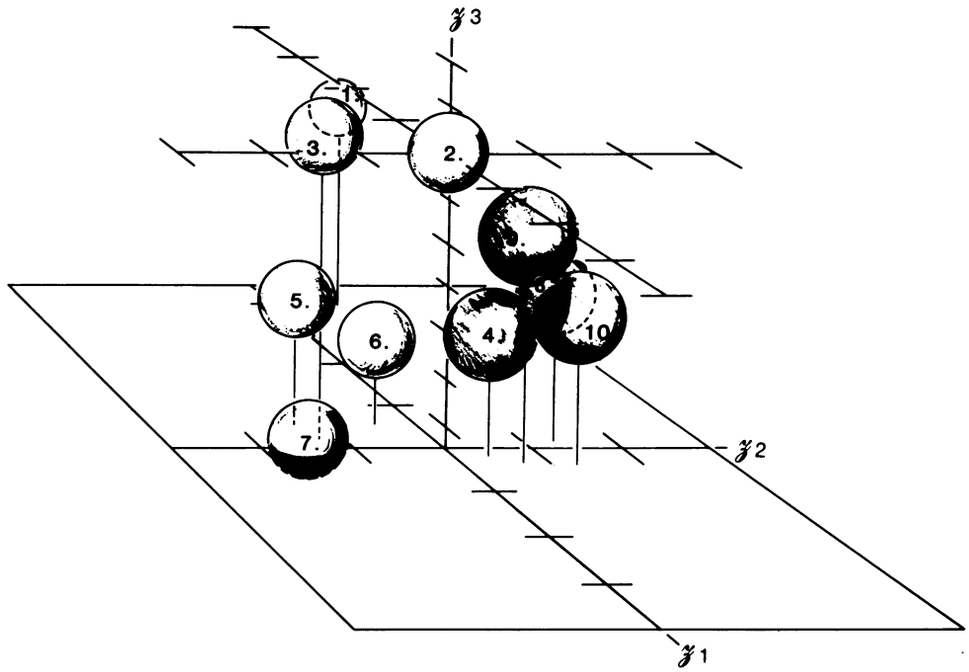

FIG. 2. Three-dimensional representation of the physiological states in terms of the first three canonical variables based on a one-way multivariate analysis of variance of 10 physiological states. The canonical variables were standardized and the means for the EU state were subtracted to center that state at 0, 0, 0. The canonical variables form the z_1, z_2, and z_3 axes. The three coordinates for each state mean were then used to plot the geometric relationships of the states. The states are identified as 1, TX; 2, EU; 3, FAST; 4, FAST + T3; 5, DB−M; 6, DB−M+I; 7, DB−S; 8, HiCHO; 9, EU + T3; and 10, HYPER as defined in Table 1.

Cluster analysis of the 19 individual translation products also clearly showed the tendencies of certain products to shift in a similar fashion in response to the various pathophysiological stimuli applied. The following primary clusters were found: (3, 8), (17, 24, 25), (14, 40), and (4, 5, 11). The remaining products were distinct.

An analysis of the canonical variables showed that 90% of the among-state variation encountered in this study could be explained by the first three canonical variables, thus indicating that relatively little information is lost if attention is restricted to the first three variables. Although the standard statistical test of dimensionality (15) indicated more canonical variables were required for a complete summary, our finding that three dimensions accounted for 90% of the observed among-state variation facilitated a geometric representation of the relationship of individual states in a three-dimensional plot (Fig. 2). The relative proximity of individual states as determined by the three-dimensional analysis was analogous to that obtained by cluster analysis. The states that merged to form larger clusters were found to be located near each other in this analysis. The relative positions of the individual states in this plot should reflect the mutual similarities and dissimilarities of these states. The close proximity of the fasted and the diabetic states indicate, as one might have anticipated, a strong similarity in the constituitive cellular reactions.

The transition from one state to another may be represented in the three-dimensional plot by a vector from one state mean to another. Two different transitions might be considered similar if the two associated vectors were parallel. For example, the vectors from EU to EU-HiCHO and from EU to HYPER are close to parallel, indicating similarity of the changes in gel pattern associated with these transitions. Similarly, vectors should be oppositely directed if two pathophysiological stimuli cause oppositely directed responses. On the other hand, one might anticipate that the angles between vectors associated with unrelated transitions would differ substantially from 0° and 180°.

DISCUSSION

Of considerable importance in this study was the finding that, by several statistical techniques, every state could be differentiated from every other state. This could not have been anticipated *a priori*, especially because the stimuli in several instances differed only in a relatively subtle fashion from each other. Thus, after the injection of a single dose of T3 into euthyroid animals, the mRNA activity profile generated differed significantly both from that characterizing the euthyroid state and the hyperthyroid state achieved after 10 successive daily doses of T_3. As illustrated in Fig. 2, the mRNA activity profile of the 1-day injected animals was situated in an intermediate position.

In a few instances, overlap was found by some statistical techniques. Thus, although by a paired comparison animals rendered severely diabetic (blood sugar > 350 mg/dl) by streptozotocin differed significantly in their genomic profiles from animals rendered not as severely diabetic (blood sugar < 250 mg/dl), discriminant and cluster analysis found enough overlap to preclude completely reliable separation. Because only two animals were used in the severely diabetic group, these results are not surprising. Therefore, we can infer from these findings that the genomic response patterns are highly sensitive to the stimuli which impinge on the organism.

It is apparent, however, that the 19 sequences that were analyzed in this study are only a minority of the 10,000–15,000 mRNA species that are expressed by the hepatic genome (16). Moreover, the identity of only a few mRNA products in these profiles is known. Previous studies indicate that spot 2 represents malic enzyme (6) and spot 22 represents glucose-6-phosphate dehydrogenase (10). As additional specific antibodies become available, it should be possible eventually to identify a substantial number of the individual sequences. Of particular interest to us was the determination of sets of spots that responded to the various stimuli in a similar fashion. Such clusters may well reflect

coordinate biological regulation. Clearly, the level of translational activity of a given mRNA is determined by multiple factors, including the rates of transcription, processing, and turnover of the mRNA, and theoretically, by the efficiency of *in vitro* translation.

The three-dimensional representation of the 10 states analyzed in this study displays several anticipated and several unanticipated relationships. The close spatial relationship of the diabetic and starved states and the small angular separation of their transition vectors from the EU state is consonant with well-established biochemical similarities between diabetes and starvation. Gluconeogenesis and lipolysis are increased and lipogenesis are diminished in both. Both states are characterized by diminished glucose utilization and an elevated level of glucagon. Nevertheless, diabetes and starvation are not identical entities, and their differences are reflected in their mRNA activity profiles. Of unusual interest to us was the finding that the presence of substantial quantitites of product 33 appeared virtually specific for the diabetic state.** Also anticipated from previous studies in our laboratory (10, 17) was the similarity between mRNA activity profiles of the hyperthyroid and high carbohydrate states. In both, there is a known increase in lipogenic enzymes. Other undefined similarities may contribute to their juxtaposition in the three-dimensional display.

At the same time, we were surprised by the marked deviation of hypothyroidism and starvation, especially because in previous studies (4) we had noted a high degree of overlap between these states in the directional changes of individual mRNA sequences; however our previous analysis did not take into account the quantitative aspects of individual shifts, but was confined simply to an analysis of the direction of such changes.

Although the data analyzed in this study could be substantially described by a three-dimensional structure (three canonical variables), it is not clear whether or not selection of additional stimuli would require additional dimensions to explain the observed variation. Even if a full description of the biological system requires more than the three dimensions derived in the present analyses, the general principles of multivariate analysis should be applicable. Our approach facilitates an objective analysis of a large data base presented by two-dimensional mRNA activity profiles and takes into account inherent biological and technical variation. Lastly, the method provides new insights into the relationship between hormonal, genetic, and pathophysiological stimuli at a pretranslational level.

**As indicated in Table 2, insulin treatment only partially reversed the alterations induced by the diabetic state. This could have been due to hepatocyte effects of streptocytocin treatment, to damage of non-beta-cell constituents of the rat pancreas, or to the nonphysiologic route of delivery of insulin into the systemic system rather than the circulation.

We thank Ms. Ana Martinez and Mr. Robert Gunville for their technical assistance and Ms. Betty Sokolosky and Ms. Jennifer Steinert for their assistance in the preparation of this manuscript. This work was supported by National Institutes of Health Grant AM19812 (to J.H.O.), a grant from the Minnesota Medical Foundation (to C.N.M.), the American Diabetes Association (to C.N.M.), Clinical Investigator Award AM00800 (to C.N.M.), and National Research Service Award AM06607 (to F.E.C.).

1. Iynedsian, P. B. & Jacot, M. M. (1981) *J. Biol. Chem.* **256,** 7068–7076.
2. Morrison, M. R., Pardue, S. & Griffin, W. S. T. (1981) *J. Biol. Chem.* **256,** 3550–3556.
3. Seelig, S., Liaw, C., Towle, H. C. & Oppenheimer, J. H. (1981) *Proc. Natl. Acad. Sci. U.S.A.* **78,** 4733–4737.
4. Carr, F. E., Seelig, S., Mariash, C. N., Schwartz, H. L. & Oppenheimer, J. H. (1983) *J. Clin. Invest.* **72,** 154–163.
5. Mariash, C. N., Seelig, S. & Oppenheimer, J. H. (1982) *Anal. Biochem.* **121,** 388–394.
6. Towle, H. C., Mariash, C. N. & Oppenheimer, J. H. (1980) *Biochemistry* **19,** 579–585.
7. Oppenheimer, J. H. & Schwartz, H. L. (1980) *Endocrinology* **107,** 1460–1468.
8. O'Farrell, P. H. (1975) *J. Biol. Chem.* **250,** 4007–4021.
9. Seelig, S., Jump, D. B., Towle, H. C., Liaw, C., Mariash, C. N., Schwartz, H. L. & Oppenheimer, J. H. (1982) *Endocrinology* **110,** 671–673.
10. Liaw, C., Seelig, S., Mariash, C. N., Oppenheimer, J. H. & Towle, H. C. (1983) *Biochemistry* **22,** 213–221.
11. Bliss, C. I. (1965) *Statistics in Biology* (McGraw-Hill, New York), Vol. 1, pp. 108–110.
12. Morrison, D. F. (1976) *Multivariate Statistical Methods* (McGraw-Hill, New York), 27th Ed., pp. 33–34, 136–141, 180.
13. Dunnet, C. W. (1980) *J. Am. Stat. Assoc.* **75,** 789–795.
14. Anderberg, M. R. (1973) *Cluster Analysis for Applications* (Academic Press, New York).
15. Rao, C. R. (1973) *Linear Statistical Inference and Its Applications* (Wiley, New York), 2nd Ed., p. 559.
16. Savage, M. J., Sala-Trepat, J. M. & Bonner, J. (1978) *Biochemistry* **17,** 460–467.
17. Mariash, C. N., McSwigan, C. R., Towle, H. C., Schwartz, H. L. & Oppenheimer, J. H. (1981) *J. Clin. Invest.* **68,** 1485–1490.